

57672

From: Davis, Natalie
Sent: Monday, January 07, 2002 11:12 AM
To: STIC-Biotech/ChemLib
Subject: 09/823101

Please search SEQ ID NO: 1-3

Natalie A. Davis, PhD
Patent Examiner
Art Unit 1642
CM1, Rm 8B13
Mailbox 8E12
Ph (703) 308-6410

CRFE

Point of Contact:
Mona Smith
Technical Info. Specialist
CM1 12C14 Tel: 308-3278

Searcher: M. Smith
Phone: _____
Location: _____
Date Picked Up: 1/7/02
Date Completed: 1/6
Searcher Prep/Review: 10
Clerical: _____
Online time: 10

TYPE OF SEARCH:
NA Sequences: 3
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2002, 08:44:55 ; Search time 3002.62 seconds
(without alignments)
841.018 Million cell updates/sec

Title: US-09-823-101-2

Perfect score: 235

Sequence: 1 acctctaccactgctgac.....ttcttagcaagtgagctt 235

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estom:*
- 5: em_estpl:*
- 6: em_estba:*
- 7: em_estro:*
- 8: em_estov:*
- 9: em_hic:*
- 10: gb_est1:*
- 11: gb_est2:*
- 12: gb_hic:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pln:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	49.4	21.0	511	13	AZ761057	AZ761057 1M0555K12
2	39.8	16.9	432	13	BH078711	BH078711 RPCI-24-3
3	38.6	16.4	578	13	AZ376465	AZ376465 1M0130P16
4	38.6	16.4	591	13	AZ375783	AZ375783 1M0129P16
5	38.2	16.3	606	13	BH060091	BH060091 RPCI-24-3
6	38.2	16.2	417	10	BE253748	BE253748 601112353
7	37.8	16.1	469	13	AQ655258	AQ655258 Sheared D
8	37.8	16.1	651	13	AQ643656	AQ643656 RPCI93-Dp
9	37.8	16.1	758	11	BG583874	BG583874 102403800
10	37.6	16.0	863	13	CNS012SP	AL101971 Drosophil
c 11	37.6	16.0	1101	13	CNS007IW	AL067324 Drosophil
12	37.4	15.9	522	11	BG284982	BG284982 602409245

c 13	36.8	15.7	560	13	AZ429688	AZ429688 1M0213H13
14	36.4	15.5	422	10	AI653440	AI653440 tq94b07.x
c 15	36.4	15.5	496	13	AZ128979	AZ128979 OSJNB006
16	36.2	15.4	409	11	H60465	H60465 yr42a02.r1
c 17	36.2	15.4	841	10	AL569885	AL569885 AL569885
18	36.2	15.4	845	11	BF578564	BF578564 602093265
c 19	36.2	15.4	923	13	CNS04NPT	AL299018 Tetraodon
20	35.8	15.3	1101	13	CNS002B1	AL097447 Drosophil
c 21	35.6	15.1	1340	11	BE890770	BE890770 601431210
22	35.4	15.1	786	13	CNS02LRS	AL203185 Tetraodon
c 23	35.4	15.1	997	13	CNS005TE	AL060767 Drosophil
24	35.4	15.1	1101	13	CNS006C9	AL072006 Drosophil
c 25	35.2	15.0	439	10	AI879059	AI879059 au53h09.y
26	35.2	15.0	482	10	AW403656	AW403656 UI-HF-BK0
c 27	35.2	15.0	622	10	AW953270	AW953270 EST365340
28	35.2	15.0	629	11	BG684818	BG684818 602636533
c 29	35.2	15.0	780	11	BF128982	BF128982 601811294
30	35.2	15.0	835	11	BF337210	BF337210 602035016
c 31	35.2	14.9	424	13	AZ496846	AZ496846 1M0333N23
32	34.8	14.8	291	10	BB528060	BB528060 BB528060
c 33	34.8	14.8	501	13	AZ193303	AZ193303 SP_1023.A
34	34.8	14.8	873	13	AZ128128	AZ128128 OSJNB008
c 35	34.8	14.8	984	11	BG497837	BG497837 602543127
36	34.6	14.7	493	10	AI399104	AI399104 NCW09A6T3
c 37	34.6	14.7	690	11	BI197022	BI197022 602756093
38	34.6	14.7	783	11	BG854194	BG854194 1024039C0
c 39	34.6	14.7	813	11	BF684008	BF684008 602141341
40	34.6	14.7	1100	13	CNS05P6E	AL329423 Tetraodon
c 41	34.4	14.6	609	13	AQ717767	AQ717767 HS_5507.B
42	34.4	14.6	879	11	BG778671	BG778671 602667947
c 43	34.4	14.6	987	13	CNS00418	AL066537 Drosophil
44	34.4	14.6	1652	11	BF863487	BF863487 963044C08
c 45	34.2	14.6	484	13	AQ198109	AQ198109 RPCI11-60

ALIGNMENTS

RESULT 1	AZ761057/c	511 bp	DNA	GSS	16-FEB-2001
LOCUS	1M0555K12F	Mouse 10kb plasmid	UUGC1M library	Mus musculus genomic	
DEFINITION	clone UUGC1M0555K12 F, DNA sequence.				
ACCESSION	AZ761057.1	GI:12869564			
VERSION	AZ761057.1				
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 511)				
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhauser,A. and Wright,D.,Weiss,R.				
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0555 row: K column: 12 Seq primer: CGTTGTAACGACGCCAGT Class: plasmid ends High quality sequence stop: 511. Location/Qualifiers 1. .511				

us-09-823-101-2.rst

Tue Jan '8 10:26:12 2002

```

source
1. .578
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0130P16"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 156 a 119 c 160 g 143 t
ORIGIN

Query Match 16.4%; Score 38.6; DB 13; Length 578;
Best Local Similarity 56.8%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 54;

QY . 104 aagggagatccaaacccctggagaagacagcagtagagtggtggcgagtgaggagcag 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 AAGGAGCGAACCAATCCCTTAATGATTAAGAGCCATCACTGGCGAGTAGGATCGGA 455

QY . 164 gaaaggttggtgacagatcaggaggtgtctgacctttttcttgagaaattctagg 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 456 GAGAGGAAGAGAGAGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 515

QY 224 caagt 228
||| |
Db 516 CAACT 520

RESULT 4
AZ375783 591 bp DNA GSS 02-OCT-2000
LOCUS IM0129P16F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0129P16 F, DNA sequence.
ACCESSION AZ375783
VERSION AZ375783.1 GI:10489483
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 591)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mamoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

1. .591
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0129P16"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 158 a 124 c 164 g 145 t
ORIGIN

Query Match 16.4%; Score 38.6; DB 13; Length 591;
Best Local Similarity 56.8%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 54;

QY 104 aagggagatccaaacccctggagaagacagcagtagagtggtggcgagtgaggagcag 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 AAGGAGCGAACCAATCCCTTAATGATTAAGAGCCATCACTGGCGAGTAGGATCGGA 456

QY 164 gaaaggttggtgacagatcaggaggtgtctgacctttttcttgagaaattctagg 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 457 GAGAGGAAGAGAGAGGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 516

QY 224 caagt 228
||| |
Db 517 CAACT 521

RESULT 5
BH060091 606 bp DNA GSS 18-JUL-2001
LOCUS RPCI-24-326K19.TV RPCI-24 Mus musculus genomic clone RPCI-24-326K19
DEFINITION , DNA sequence.
ACCESSION BH060091
VERSION BH060091.1 GI:14870888
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 606)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.

```

TITLE JOURNAL COMMENT

Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-326K19.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
plate: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 326 row: K column: 19
Seq primer: T7
Class: BAC ends.

FEATURES source

Location/Qualifiers
1. 606
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-326K19"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTANBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTANBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 162 a 130 c 174 g 140 t

Query Match 16.3%; Score 38.2; DB 13; Length 606;
Best Local Similarity 56.9%; Pred. No. 4;
Matches 70; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 104 aaggaggatcccaaccctggaagacacagcagtagaagtgagcagtgaggagcag 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 AAGGAGGACCAACATCCCTTAATGATATAGAGGACCATCATCTGGCGGAGTAGGCGGA 455
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 gaaagtgtggtgacagatcagggaggtgtctgacccctttcttgaggaattctttagg 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 456 GAGAGAAGAGAGAGGGGGGAGAGAGAGAGAGGGGCCCTTTTGGATGAGGACAGATGGAG 515
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 224 caa 226
|||
Db 516 CAA 518

RESULT 6
BE253748
LOCUS BE253748 417 bp mRNA EST 13-JUL-2000
DEFINITION 601112353F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3353172 5',
mRNA sequence.
ACCESSION BE253748
VERSION BE253748.1 GI:9124168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 417)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Scrausberg, Ph.D.
Email: cga@ps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM153 row: i column: 13
High quality sequence stop: 410.

FEATURES source

Location/Qualifiers
1. 417
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3353172"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 110 a 103 c 136 g 68 t
ORIGIN

Query Match 16.2%; Score 38; DB 10; Length 417;
Best Local Similarity 51.1%; Pred. No. 4.2;
Matches 89; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 41 agagagtagagagatcagagtcagctctcatagccaaataactatgcttaccatgccg 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 AAGAGCGAGAAATGGAAGAGAGCGCCCTGCTAACGGGAATGCTAATGAGGAATGGG 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 101 aggaaggggagatcccaaccctggaagacacagcagtagaagtgaggcagtgaggagg 160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 AGCAGGAGCTGACATGAGGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 161 caggaaaggttggtgacagatcagggaggtgtctgacccctttcttgaggaa 214
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 AGGAGAAGATGGTGTGGGTGCGAGGAGAGAGGATGAGATGAGATGAGGAA 396
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AO655258
LOCUS AO655258 469 bp DNA GSS 22-JUN-1999
DEFINITION Sheared DNA-20M16.Tf Sheared DNA Trypanosoma brucei genomic clone
AO655258
ACCESSION AO655258
VERSION AO655258.1 GI:5148444
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 469)
AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-20M16.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:

Qy	89	cttacagtcctccgaggaaggagggagatcccaaaccttggaagacagcgagtagaagtgg	148
Db	533	MCACCCGAAACAGGTTWAGGCTAGTGAAGACAGAGAGTGCACACGCGCCGAAAGAGTGG	474
Qy	149	gcagtgaggagcgagcaaaagtgtgggtg	176
Db	473	TCGSGCGGAYGCANAGRCGGGAGGGTG	446
RESULT	12		
LOCUS	BC284982		
DEFINITION	602409245F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4538219	5'	21-FEB-2001
ACCESSION	BC284982		
VERSION	BC284982		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCTD/DTP cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10454 row: j column: 12 High quality sequence stop: 522.		
FEATURES	Location/Qualifiers		
source	1. .522		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4538219"		
	/clone_lib="NIH_MGC_91"		
	/tissue_type="adenocarcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.4 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	159 a 123 c 161 g 79 t		
ORIGIN			
Query Match	15.9%;	Score 37.4;	DB 11; Length 522;
Best Local Similarity	53.0%;	Pred. No. 6.4;	
Matches	80; Conservative	0; Mismatches	71; Indels 0; Gaps 0;
Qy	41	aggaggtaggagatgaagtcaggctctcatagcacaatactatgcttacagtcctcg	100
Db	241	AAGAGGCAGAAAATGGAAGAGACGCCCTCTACTACGGGAATGCTAATGAGAAAATGGG	300
Qy	101	aggaaaggagatcccaaaccttggaagacagcgagtagaagtgggcgagtgaggag	160
Db	301	AGCAGGAGGCTGACAAATCAGGTAGACCAAGAGAGAGAGAGGTGGGAGGAGAGAGG	360
Qy	161	caggaaaggttggtgcacatcaggagaggg	191
Db	361	AGGAAGAAGAACGCTGACTGCTGTGAGGAAGAG	391
RESULT	13		
LOCUS	AZ429688/c		
DEFINITION	1M0213H13R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0213H13 R, DNA sequence.		
ACCESSION	AZ429688		
VERSION	AZ429688.1		
KEYWORDS	GSS.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.		
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0213 row: H column: 13 Seq primer: CACACAGAAACAGCTATGACC Class: plasmid ends High quality sequence stop: 560.		
FEATURES	Location/Qualifiers		
source	1. .560		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="UUGCLM0213H13"		
	/clone_lib="Mouse 10kb plasmid UUGCLM library"		
	/sex="Male"		
	/lab_host="E. Coli strain XL10-Gold, TI-resistant, P-"		
	/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114 gb AF129072.1), a copy-number inducible derivative of plasmid RL1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."		
BASE COUNT	23 a 274 c 15 g 248 t		
ORIGIN			
Query Match	15.7%;	Score 36.8;	DB 13; Length 560;
Best Local Similarity	53.5%;	Pred. No. 9.3;	
Matches	77; Conservative	0; Mismatches	67; Indels 0; Gaps 0;
Qy	43	gaggttagagagataagtcaggctctctatagccaaatactatggcttacagtcctcgag	102
Db	482	GAGGGAAGCTGAAGGGAGAAAGAGAGTGAAGAGAGAAGATGGAGGACAGAGAGAG	423
Qy	103	gaaggggagatcccaaaccttggaagacagcgagtagaagtgggcgagtgaggagca	162
Db	422	GAAGGAAGAATTCAGGCAGGCAAAATGCGGAGGAAGGAGGAGGAGAGGAGGAGGA	363

QY 203 ttcc 206
|||
Db 192 ATTC 189

Search completed: January 8, 2002, 08:45:00
Job time: 5607 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2002, 08:47:16 ; Search time 116.06 Seconds
(without alignments)
458.576 Million cell updates/sec

Title: US-09-823-101-2
Perfect score: 235
Sequence: 1 acctctaccactgctgtac.....ttcttagcgaagtgaagctt 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	43.8	18.6	7218	1	US-08-232-463-14
C 2	34.8	14.8	3482	3	US-09-111-085-3
C 3	34	14.5	2184	4	US-09-307-265A-2
C 4	31.8	13.5	289	4	US-09-007-005-17
C 5	31.8	13.5	289	4	US-09-244-796-17
C 6	31.4	13.4	2061	2	US-08-835-170-1
C 7	31.4	13.4	2061	4	US-09-359-257-1
C 8	31.4	13.4	2327	2	US-08-835-170-3
C 9	31.4	13.4	2327	4	US-09-359-257-3
C 10	31	13.2	6171	1	US-08-459-568-1
C 11	31	13.2	6171	2	US-08-399-411-1
C 12	31	13.2	6171	3	US-08-516-859A-1
C 13	30.8	13.1	2188	1	US-07-865-662F-10
C 14	30.8	13.1	2188	4	US-08-374-219B-10
C 15	30.2	12.9	1461	1	US-08-587-670A-1
C 16	30.2	12.9	1461	4	US-09-061-674-1
C 17	30.2	12.9	80246	4	US-09-078-294-4
C 18	30.2	12.9	80595	4	US-09-078-294-3
C 19	29.8	12.7	2580	3	US-09-050-863-2
C 20	29.8	12.7	5452	2	US-09-130-114-1
C 21	29.8	12.7	9500	4	US-08-910-114-1
C 22	29.8	12.7	10596	1	US-07-884-811-15
C 23	29.8	12.7	10596	1	US-07-885-971-15
C 24	29.8	12.7	10596	1	US-08-087-783A-15
C 25	29.8	12.7	10596	1	US-08-194-088B-15
C 26	29.8	12.7	10596	2	US-08-194-087-15
C 27	29.8	12.7	10596	5	PCT-US93-04648-15

C 28	29.6	12.6	1200	3	US-08-617-860B-4	Sequence 4, Appl1
C 29	29.6	12.6	1931	2	US-09-130-114-2	Sequence 2, Appl1
C 30	29.4	12.5	19056	4	US-09-272-032-8	Sequence 8, Appl1
C 31	29.2	12.4	197	1	US-08-392-678-5	Sequence 5, Appl1
C 32	29.2	12.4	197	1	US-08-457-304A-5	Sequence 5, Appl1
C 33	29.2	12.4	197	1	US-08-456-701A-5	Sequence 2, Appl1
C 34	29	12.3	5361	4	US-08-973-462-1	Sequence 1, Appl1
C 35	29	12.3	6152	4	US-08-973-462-2	Sequence 2, Appl1
C 36	28.6	12.2	1540	2	US-08-943-600A-2	Sequence 11, Appl1
C 37	28.4	12.1	325	2	US-08-332-766A-11	Patent No. 5340934
C 38	28.4	12.1	975	6	5340934-10	Sequence 1, Appl1
C 39	28.4	12.1	2089	1	US-08-552-142A-1	Sequence 1, Appl1
C 40	28.4	12.1	2089	1	US-08-910-973-1	Sequence 1, Appl1
C 41	28.4	12.1	2089	5	PCT-US95-05741-1	Sequence 1, Appl1
C 42	28.4	12.1	35524	3	US-08-923-137-1	Sequence 25, Appl1
C 43	28.2	12.0	421	2	US-08-332-766A-25	Sequence 1, Appl1
C 44	28.2	12.0	1275	2	US-08-958-642-1	Sequence 3, Appl1
C 45	28.2	12.0	1275	3	US-08-778-394-3	

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14


```
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      13.5%; Score 31.8; DB 4; Length 289;
Best Local Similarity 1.7%; Pred. No. 0.23;
Matches 3; Conservative 84; Mismatches 87; Indels 0; Gaps 0;

QY 17 tgacttggtgagtggtggtccaaaggagtaggagagataagagtcaggctctcatagc 76
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 utarururacraruracraruracraruracraruracraruracraruracrar 83

QY 77 caaatactatgcttcagtcagccgaggaaggagagagagagagagagagagagagag 136
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 84 nrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrn 143

QY 137 cagtagaagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 nrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrnrsnrn 197

RESULT 6
US-08-835-170-1
; Sequence 1, Application US/08835170
; Patent No. 5965420
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; TITLE OF INVENTION: Human Protein Kinases HYAK3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,170
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50004

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-835-170-1

Query Match      13.4%; Score 31.4; DB 2; Length 2061;
Best Local Similarity 51.8%; Pred. No. 0.65;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 93 cagtcctccgaggaaggagagagagagagagagagagagagagagagagagagagag 152
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1514 CAGGCAGATGGGAGGCTTGCTTGTGGGGGTCGCTCAGTAGGGTAAAAACGGGGT 1573

QY 153 gtggggagggcagaaaggttgggtgacagatcaggggggtgtctgacctttttcttgagg 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1574 CCCCCAGGCAGCAAGACTGGGGGACAGCACTGAAAGGGTGTGATGACTACTTGTATTATA 1633

QY 213 aaattcttaggcaagtgtg 229
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1634 GAGTCTCTTGAAGAGGTG 1650

RESULT 7
US-09-359-257-1
; Sequence 1, Application US/09359257
; Patent No. 6165766
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Xie, Wei
; TITLE OF INVENTION: HUMAN PROTEIN KINASES HYAK3
; FILE REFERENCE: GH50004X1D1
; CURRENT APPLICATION NUMBER: US/09/359,257
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: 08/835,170
; EARLIER FILING DATE: 1997-04-07
; EARLIER APPLICATION NUMBER: 60/040618
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Human
US-09-359-257-1

Query Match      13.4%; Score 31.4; DB 4; Length 2061;
Best Local Similarity 51.8%; Pred. No. 0.65;
Matches 71; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 93 cagtcctccgaggaaggagagagagagagagagagagagagagagagagagagagag 152
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1514 caggcagatgggagaggttgtctgtgggggtcgctcagctaggggtataaaagcggggt 1573

QY 153 gtggggagggcagaaaggttgggtgacagatcaggggggtgtctgacctttttcttgagg 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1574 cccccaggcagcaagactgggggacagcactgaaaggggtgtgatgactactgtttata 1633

QY 213 aaattcttaggcaagtgtg 229
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1634 gagttcttgaagaggtgtg 1650

RESULT 8
```



```
; LOCATION: 157..5275
US-08-459-568-1

Query Match      13.2%; Score 31; DB 1; Length 6171;
Best Local Similarity 62.0%; Pred. No. 1.3;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 100 gaggaagggagatccaaaccctggaagacaagcagtagaagtggcgagtgaggag 159
| ||| ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Db 3350 GGGAGGGGGAGGTGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3291

QY 160 gcaggaaggttgggtgac 178
| || | ||||| |||
Db 3290 GAAGAAAGTGTGGGGGAC 3272

RESULT 11
US-08-399-411-1/c
; Sequence 1, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 157..5275
; US-08-399-411-1

Query Match      13.2%; Score 31; DB 2; Length 6171;
Best Local Similarity 62.0%; Pred. No. 1.3;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 100 gaggaagggagatccaaaccctggaagacaagcagtagaagtggcgagtgaggag 159
| ||| ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Db 3350 GGGAGGGGGAGGTGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3291

QY 160 gcaggaaggttgggtgac 178
| || | ||||| |||
Db 3290 GAAGAAAGTGTGGGGGAC 3272

RESULT 12
US-08-516-859A-1/c
; Sequence 1, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 157..5275
; US-08-516-859A-1

Query Match      13.2%; Score 31; DB 3; Length 6171;
Best Local Similarity 62.0%; Pred. No. 1.3;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 100 gaggaagggagatccaaaccctggaagacaagcagtagaagtggcgagtgaggag 159
| ||| ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Db 3350 GGGAGGGGGAGGTGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3291

QY 160 gcaggaaggttgggtgac 178
| || | ||||| |||
Db 3290 GAAGAAAGTGTGGGGGAC 3272

RESULT 13
US-07-865-662F-10/c
; Sequence 10, Application US/07865662F
; Patent No. 5451670
; GENERAL INFORMATION:
; APPLICANT: Marcia M. Miller
; TITLE OF INVENTION: Restriction Fragment Length
; POLYMORPHISM TEST FOR HAPLOTYPING DOMESTICATED FOWL
; NUMBER OF SEQUENCES: 14
```


STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,670A
FILING DATE: 17-JAN-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34,053
REFERENCE/DOCKET NUMBER: 07949/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1461 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-587-670A-1

Query Match 12.9%; Score 30.2; DB 1; Length 1461;
Best Local Similarity 56.6%; Pred. No. 1.4;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 79 aatactatggttacagtcgccgaggaagggagagatccaaacccctggaagaagaagga 138
Db 158 AATTTCATGTTGTCCTGAAGAAGATGACTTTGTATGGGAATCCATGGGAAGAAGCATGAGG 99
QY 139 gtagaagtggcgagtggtggcagcaggaaggttggtga 177
Db 98 TGAGAAGCCAGGTGAGGGCGGCGAGGAGGCTGGGTCA 60

Search completed: January 8, 2002, 08:47:28
Job time: 5680 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2002, 08:53:02 ; Search time 279.63 Seconds
(without alignments)
720.493 Million cell updates/sec

Title: US-09-823-101-2
Perfect score: 235
Sequence: 1 accttcaccactgctgac.....ttcttagcgaagtgaagctt 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_1101.*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35.2	15.0	1099	21	Human ORFX ORF2753
C 2	34.8	14.8	1974	18	Porcine retrovirus
C 3	34.8	14.8	3482	20	Pig endogenous ret
C 4	34.8	14.8	8196	18	Porcine retrovirus
C 5	34.8	14.8	8209	18	Porcine retrovirus
C 6	34.6	14.7	500	21	Cat flea head and
C 7	34.4	14.6	626	22	Probe #12056 for g
C 8	34.4	14.6	626	22	Probe #16105 used
C 9	34.4	14.6	626	22	Probe #7814 used t
C 10	34.4	14.6	2784	18	Luciferase gene un
C 11	34	14.5	1954	22	Human polynucleoti

12	34	14.5	2182	22	AAI59351	Human polynucleoti
13	34	14.5	2184	22	AAD06295	Human Ras suppress
14	34	14.5	3251	22	AAH16802	Human CDNA sequenc
15	34	14.5	3253	22	AAH15878	Human CDNA sequenc
c 16	33.6	14.3	1297	21	AAA26368	Human secreted pro
17	33.2	14.1	575	22	AAI23315	Probe #13248 for g
18	33.2	14.1	575	22	AAI48636	Probe #17322 used
19	33.2	14.1	575	22	AAI08956	Probe #8947 used t
20	33.2	14.1	1969	22	AAI14105	Probe #4038 for ge
21	33.2	14.1	1969	22	AAI35486	Probe #4172 used t
22	33.2	14.1	1969	22	AAI03958	Probe #3949 used t
c 23	32.8	14.0	484	22	AAI38784	Human secreted pro
24	32.8	14.0	2043	19	AAV59728	Human protein enco
25	32.8	14.0	3149	22	AAH99797	Human Th2-specific
c 26	32.8	14.0	3868	21	AAAO9058	Human Th2-specific
c 27	32.8	14.0	4421	21	AACT6287	Human ORFX ORF1842
c 28	32	13.6	1535	21	AAZ93355	Sequence encoding
29	31.6	13.4	2800	22	AAD07131	Canine retinitis p
30	31.6	13.4	2803	22	AAD07132	Canine retinitis p
31	31.6	13.4	2805	22	AAD07130	Canine retinitis p
32	31.4	13.4	964	20	AAZ97739	Extended human sec
33	31.4	13.4	1221	21	AAZ37678	Arabidopsis thalia
34	31.4	13.4	1671	22	AAH99858	Human protein enco
35	31.4	13.4	1892	20	AAZ24890	Human secreted pro
36	31.4	13.4	1895	22	AAD08658	Human cancer assoc
37	31.4	13.4	1898	22	AAZ30056	Human cDNA encodin
38	31.4	13.4	1899	20	AAZ06785	Human adult testis
39	31.4	13.4	1899	22	AAZ91465	Human PRO306 cDNA
40	31.4	13.4	1912	21	AAZ50927	Human Protease and
41	31.4	13.4	2061	19	AAZ57439	Human protein kina
42	31.4	13.4	2061	21	AAA47744	HYAK3-2 alpha codi
43	31.4	13.4	2061	22	AAD06042	Human YAK3a (HYAK3
44	31.4	13.4	2140	21	AAZ46161	cDNA sequence enco
45	31.4	13.4	2141	22	AAZ44689	Novel protein kina

ALIGNMENTS

RESULT 1
AAC77198/c
ID AAC77198 standard; cDNA; 1099 BP.
XX
AC AAC77198;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2753 polynucleotide sequence SEQ ID NO:5505.
XX

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antinflammatory;
antifungal; antibacterial; antifungal; antineoplastic; antithyroid;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antinflammatory disease; coagulation;
thrombosis; contraceptive; ss.

XX Homo sapiens.
OS WO200058473-A2.
XX
PN
XX
XX PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR

PN W09853104-A2.
 XX
 PD 26-NOV-1998.
 XX
 PF 18-MAY-1998; 98WO-GB01428.
 XX
 PR 16-MAY-1997; 97GB-0010154.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 XX
 XX Stoye JP, Weiss RA;
 XX
 DR WPI; 1999-045324/04.
 DR P-PSDB; AAW85452.
 XX
 PT Newly isolated nucleic acid probe capable of hybridising to either
 PT the PERV-A or PERV-B env gene - useful in the detection of
 PT retroviruses, and their subtypes, in a sample of porcine/human
 PT tissue
 XX
 XX Claim 3; Page 21-23; 36pp; English.
 XX
 CC The present sequence encodes a pig endogenous retrovirus (PERV)-B
 CC envelope protein. PERV exists in two different subtypes, PERV-A and
 CC PERV-B. The differences are reflected in sequence divergence in the
 CC envelope genes. Probes and primers can be derived from the envelope
 CC (env) genes of PERV-A and PERV-B. The probes and primers are used in
 CC a method to detect retroviruses in a sample of porcine/human tissue,
 CC particularly primary porcine tissue and human cell lines that have been
 CC cultivated in the presence of a porcine cell line, or human tissue from
 CC a patient with a xenotransplant. Subtype of PERV in a sample containing
 CC one of the PERV env genes can also be determined.
 XX
 XX Sequence 3482 BP; 927 A; 854 C; 867 G; 834 T; 0 other;
 SQ

Query Match 14.8%; Score 34.8; DB 20; Length 3482;
 Best Local Similarity 55.9%; Pred. No. 0.38;
 Matches 66; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 Qy 106 ggggagatcaaacccctggagaagacagcagtagaagtggcgagtgaggcgagga 165
 Db 1896 GAGGATAGACAAGCAACCAAGAGAGAGTGGCAGGTGGTGGTGGTGGTGGTGG 1837
 Qy 166 aagttgggtgacagatcagggaggtgtctgacattttcttgagaaattcttagg 223
 Db 1836 AAGCTCCCTGGATGAGACTGAGAGTCTCTGCTCTCTTAACAGGAAACACCTGGGG 1779

RESULT 4
 AAV09699/C
 ID AAV09699 standard; DNA; 8196 BP.
 XX
 AC AAV09699;
 XX
 DT 19-MAY-1998 (first entry)
 XX
 DE Porcine retrovirus DNA encoding, GAG, POL and ENV.
 XX
 KW Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein;
 KW vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
 XX
 OS Porcine retrovirus.
 XX
 FH Key Location/Qualifiers
 FT CDS 576..2126
 FT /*tag= a
 FT /product= GAG protein
 FT /note= "virion core polypeptide"
 FT CDS 2143..5733
 FT /*tag= b
 FT /product= POL protein
 FT /note= "polymerase peptide sequence as given in

FT CDS 5606..7576 Specification"
 FT /*tag= c
 FT /product= ENV protein
 FT /note= "envelope protein"
 XX
 PN W09740167-A1.
 XX
 PD 30-OCT-1997.
 XX
 PF 18-APR-1997; 97WO-GB01087.
 XX
 PR 10-FEB-1997; 97GB-0002668.
 PR 19-APR-1996; 96GB-0008164.
 XX
 XX (IMUT-) IMUTRAN LTD.
 PA (QONE-) Q-ONE BIOTECH LTD.
 XX
 PI Galbraith DN, Haworth C, Lees GM, Smith KT;
 XX
 XX WPI; 1997-535851/49.
 XX
 PT Polynucleotide encoding porcine retrovirus expression product -
 PT useful to develop products for use in vaccines, diagnosis and
 PT xeno-transplantation
 XX
 PS Claim 4; Fig 2; 69pp; English.
 XX
 CC This DNA sequence encodes the porcine retrovirus (PoEV) virion core
 CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
 CC proteins can be used to develop viral vaccines, antisense nucleic acids,
 CC ribozymes and other antiviral agents. They can also be used in
 CC xeno-transplantation technology and as diagnostic tools.
 XX
 XX Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T; 3 other;
 SQ

Query Match 14.8%; Score 34.8; DB 18; Length 8196;
 Best Local Similarity 55.9%; Pred. No. 0.51;
 Matches 66; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 Qy 106 ggggagatcaaacccctggagaagacagcagtagaagtggcgagtgaggcgagga 165
 Db 6588 GAGGATAGACAAGCAACCAAGAGAGTGGCAGGTGGTGGTGGTGGTGGTGG 6529
 Qy 166 aagttgggtgacagatcagggaggtgtctgacattttcttgagaaattcttagg 223
 Db 6528 AAGCTCCCTGGATGAGACTGAGAGTCTCTGCTCTCTTAACAGGAAACACCTGGGG 6471

RESULT 5
 AAV09700/C
 ID AAV09700 standard; DNA; 8209 BP.
 XX
 AC AAV09700;
 XX
 DT 20-MAY-1998 (first entry)
 XX
 DE Porcine retrovirus DNA encoding the LTR and GAG, POL and ENV proteins.
 XX
 KW Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein;
 KW virion core polypeptide; polymerase protein; envelope protein;
 KW vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
 XX
 OS Porcine retrovirus.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..61
 FT /*tag= a
 FT /note= "R LTR domain"
 FT LTR 1..588
 FT /*tag= b
 FT misc_feature 62..143


```
XX (HESK-) HESKA CORP.
XX PA
XX PI
XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX XX
XX DR WPI; 2000-656323/63.
XX XX
XX XX Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX PT acids useful for the prevention, diagnosis and treatment of flea
XX PT infestations -
XX XX
XX PS Claim 26; Page 285; 964pp; English.
XX XX
XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX CC or head and nerve cord (HNC) tissue. The invention also relates to the
XX CC encoded proteins. The invention additionally encompasses expression
XX CC constructs, recombinant viruses and recombinant cells comprising the
XX CC nucleic acids of the invention, recombinant production of the proteins,
XX CC antibodies against the proteins, a method of identifying inhibitors of
XX CC the proteins, and compositions comprising the inhibitors for
XX CC administration to an animal. The nucleic acids, and the proteins they
XX CC encode may be used in the prevention, treatment and diagnosis of diseases
XX CC associated with flea infestations. For example, the nucleic acids may be
XX CC used to produce an HMT or HNC protein according to standard recombinant
XX CC DNA methodology by inserting the nucleic acids into a host cell and
XX CC culturing the cell to express the protein. The HMT and HNC nucleic acids
XX CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX CC and quantitate the presence of cat flea or other homologous nucleic acid
XX CC sequences in samples. They may also be used to study the expression and
XX CC function of the proteins and their role in metabolism. The HMT and HNC
XX CC proteins may be used as antigens in the production of specific
XX CC antibodies, and in assays to identify modulators (agonists and
XX CC antagonists) of HMT and/or HNC protein expression and activity. The
XX CC anti-HMT/HNC protein antibodies and antagonists may also be used to
XX CC downregulate protein expression and activity. The antibodies may also be
XX CC used as diagnostic agents for detecting the presence of flea polypeptides
XX CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX CC present sequence represents a cat flea HNC cDNA of the invention.
XX XX
XX XX Sequence 500 BP; 33 A; 216 C; 23 G; 110 T; 118 other;
XX XX

Query Match 14.7%; Score 34.6; DB 21; Length 500;
Best Local Similarity 39.5%; Pred. No. 0.22;
Matches 70; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 23 ggtgagtggtgtccaggaggttaggagagataagagagtcaggtctcatagccaaata 82
DB 288 GGAANGNGGNGGAGGAGGAGGAGGAGGNGGAGGNGGNGGNGGNGGNGGNGGNGGAGGA 229

QY 83 ctatggtctacagtcctccaggagggagagatccaaacctggaagaagcaagcgtag 142
DB 228 GAGANGAGNNNNNGGAGGGGNGGAGGAGANAAAGGNGGNGGNGGNGGAGAGANGTAG 169

QY 143 aegtggcgagtgaggagcagaaagtgtggtgacagatcaggaggtgtgtgac 199
DB 168 ANGNGGGGNGGNGGAGGNGGAGGAGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGCC 112

RESULT 7
AAI22123/C
ID AAI22123 standard; DNA; 626 BP.
XX XX
XX AC AAI22123;
XX XX
XX XX 12-OCT-2001 (first entry)
XX DT
XX DE Probe #12056 for gene expression analysis in human cervical cell sample.
XX DE
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX XX

OS Homo sapiens.
XX XX
XX PN WO200157278-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00670.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX DR WPI; 2001-488901/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX XX
XX PS Claim 25; SEQ ID No 12056; 487pp; English.
XX XX
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP). The present sequence is one such probe. The SENPs are derived
XX CC from human Hela cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 626 BP; 100 A; 214 C; 104 G; 208 T; 0 other;
XX XX

Query Match 14.6%; Score 34.4; DB 22; Length 626;
Best Local Similarity 51.3%; Pred. No. 0.27;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 35 tgtccaaggaggttaggagagataagagtcagggtctcatagccaaatactatggttaca 94
DB 384 TGAACCTGGAAGGAGGAGGTTGCAGTGAGCCCAATATCATGCCACTATCTCCAGCCTGGG 325

QY 95 gtcccaggagaaggaggagatccaaacctggaagaagcaagcgtagtagagtggtgag 154
DB 324 TGNACAGGCGAGACTCCATCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265

QY 155 gggaggcaggaagggttggtgacagatcaggagg 190
DB 264 AAGAAGGGGAAGGGGAAGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 229

RESULT 8
AAI47419/C
ID AAI47419 standard; DNA; 626 BP.
XX XX
XX AC AAI47419;
XX XX
XX DT 17-OCT-2001 (first entry)
XX DE
XX DE Probe #16105 used to measure gene expression in human placenta sample.
XX DE
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX XX
```

PN WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 03-JUN-2000; 2000US-0608408.

XX 30-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0236359.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000US-0236359.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

XX Claim 25; SEQ ID No 7814; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes.

XX The present sequence is one such probe. The probes are useful for

XX measuring human gene expression in a human breast sample, where the probe

XX hybridises at high stringency to a nucleic acid expressed in the human

XX breast. The probes are useful for predicting, diagnosing, grading,

XX staging, monitoring and prognosing diseases of the human breast,

XX particularly those diseases with polygenic aetiology. The diseases

XX include: breast cancer, disorders of development, inflammatory diseases

XX of the breast, fibrocystic changes, proliferative breast disease and

XX non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 626 BP; 100 A; 214 C; 104 G; 208 T; 0 other;

XX Query Match 14.68; Score 34.4; DB 22; Length 626;

XX Best Local Similarity 51.38; Pred. No. 0.27; Indels 0; Gaps 0;

XX Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 35 tgtccaaggaggttagagagataagagtcaggctctctatagccaaatactatggtttaca 94

Db 384 TGAAGTTCGAGGAGGAGGTTGCAGTGAGCCAAATATCATGCGCACTATACTCCAGCCTGGG 325

Qy 95 gtcccgaggaaggagagatccaaacccctggagagacagcagtagaagtggtggcgagt 154

Db 324 TGACAGACGCGAGACTCCATCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265

Qy 155 gggagggcaggaaaggttggtgacagatcaggagg 190

Db 264 AAGAAGGGGAAGGGGAGGGGAG 229

RESULT 10

AAT65102/c

ID AAT65102 standard; DNA; 2784 BP.

XX AAT65102;

XX 24-FEB-1998 (first entry)

XX Luciferase gene under control of rat MLC-2 gene promoter.

XX Luciferase; upstream regulatory region; MLC-2 promoter; gene therapy;

XX rat cardiac myosin light chain gene; heart disease; cardiac muscle; ss.

XX Chimeric - Synthetic.

XX Chimeric - Rattus sp.

XX Chimeric - Adenovirus.

XX Key Location/Qualifiers

XX misc_feature 1..156

XX /*tag= a

XX /note= "encodes Adenovirus Ad5 psi encapsidation"

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 1954 BP; 439 A; 504 C; 537 G; 474 T; 0 other;

Query Match 14.5%; Score 34; DB 22; Length 1954;
 Best Local Similarity 59.2%; Pred. No. 0.54;
 Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 104 aaggggagatccaaacccctggaagaagacagcagtagaagtggcgagtgaggagcag 163

Db 194 aaggggagattcagcccggaacctggcgatggcgtggaagcgtggaacagcaggaaggaa 253

Qy 164 gaaaggttggtgacagatcagggaggggtgtctgacct 201

Db 254 aggcgtctggggagagagcgcgagctggtgtgtact 291

RESULT 12

ID AAI59351 standard; cDNA; 2182 BP.

XX AAI59351;

DT -22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1554.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM40195.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 1554; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 2182 BP; 477 A; 552 C; 587 G; 566 T; 0 other;

Query Match 14.5%; Score 34; DB 22; Length 2182;

Best Local Similarity 59.2%; Pred. No. 0.57;

Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 104 aagggagatccaaacccctggaagaagacagcagtagaagtggcgagtgaggagcag 163

Db 429 aaggggagattcagcccggaacctggcgatggcgtggaagcgtggaacagcaggaaggaa 488

Qy 164 gaaaggttggtgacagatcagggaggggtgtctgacct 201

Db 489 agcgtctggggagagagcgcgagctggtgtgtact 526

RESULT 13

AAD06295

ID AAD06295 standard; cDNA; 2184 BP.

XX AAD06295;

XX 10-AUG-2001 (first entry)

XX Human Ras suppressor sur-5 cDNA.

XX Ras suppressor; sur-5; cytostatic; Ras signalling; gene therapy;
 KW cell growth; differentiation; proliferation; transgenic animal;
 KW cancer; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

EH 1..1998

FT CDS

FT /*tag= a

FT /product= "SUR-5 protein"

FT /transl_except= (pos:1756..1758, aa:Asp)

FT /note= "CDS does not include start codon"

FT /partial

XX US6225456-B1.

XX 01-MAY-2001.

XX 06-MAY-1999; 99US-0307265.

XX 07-MAY-1998; 98US-0084590.

XX (UYTE-) UNIV TECHNOLOGY CORP.

XX Gu T, Orita S, Han M;

XX WPI; 2001-342417/36.

XX P-PSDB; AAE02345.

XX Novel Ras suppressor SUR-5 polynucleotides from human, murine,

XX Caenorhabditis elegans, for treating Ras-mediated effects, especially

XX cancer and producing SUR-5 polypeptides for screening drugs altering

XX Ras signalling -

```
XX Claim 1; Column 57-60; 50pp; English.
PS
XX The invention relates to Ras suppressor sur-5 polynucleotides and
CC polypeptides (SUR-5) from human, murine and Caenorhabditis elegans
CC (nematode). SUR-5 polynucleotides are useful for producing poly-
CC peptides by recombinant techniques, which are useful as targets for
CC screening drugs that alter Ras signalling and the physiological
CC effects of Ras such as effects on cell growth, differentiation and
CC proliferation. SUR-5 genes are useful in gene therapy for treating
CC diseases or altering physiological states characterised by unwanted
CC proliferation of cells or other Ras-mediated effects, especially
CC cancer. The invention also provides methods of producing transgenic
CC animals expressing SUR-5.
CC The present sequence is human Ras suppressor sur-5 cDNA.
XX
SQ Sequence 2184 BP; 501 A; 556 C; 643 G; 484 T; 0 other;

Query Match 14.5%; Score 34; DB 22; Length 2184;
Best Local Similarity 59.2%; Pred.No. 0.57;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 104 aagggagatccaaacccctggaagaagacaagcagtagaagtggcgagtgaggagcag 163
DB 1346 aagggagagattcagcccgaaacctggcattggccgtggaagcgtggaacgagaaggaa 1405

QY 164 gaaagttggtgacagatcagggaggggtgtctgacct 201
DB 1406 agcggtctctgggagagagcgcgagctggtgtgtact 1443

RESULT 14
AAHI5602
ID AAHI5602 standard; cDNA; 3251 BP.
XX
AC AAHI5602;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:16049.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 16049; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
```

```
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
```

SQ Sequence 3251 BP; 706 A; 866 C; 926 G; 753 T; 0 other;

```
Query Match 14.5%; Score 34; DB 22; Length 3251;
Best Local Similarity 59.2%; Pred.No. 0.65;
Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 104 aagggagatccaaacccctggaagaagacaagcagtagaagtggcgagtgaggagcag 163
DB 1514 aagggagagattcagcccgaaacctggcattggccgtggaagcgtggaacgagaaggaa 1573

QY 164 gaaagttggtgacagatcagggaggggtgtctgacct 201
DB 1574 agcggtctctgggagagagcgcgagctggtgtgtact 1611
```

```
RESULT 15
AAHI5678
ID AAHI5678 standard; cDNA; 3253 BP.
XX
AC AAHI5678;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14039.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
```

```
XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
```

PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 14039; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ .Sequence 3253 BP; 706 A; 864 C; 927 G; 756 T; 0 other;
 Query Match 14.5%; Score 34; DB 22; Length 3253;
 Best Local Similarity 59.2%; Pred. No. 0.65;
 Matches 58; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
 Qy 104 aaggggagatccaaacccctgaaagaagacagagcagtagaagtggcgagtgaggagcag 163
 |||||
 Db 1515 aaggggagattcaggcccggaacctggcgatggcgtggaagcgtggaacgaggaagaa 1574
 |||||
 Qy 164 gaaaggttgggtacagatcaggaggggtgtctgaacct 201
 |||||
 Db 1575 aggcggtctgggagagagagcgcgagctggtgtgtact 1612
 |||||
 Search completed: January 8, 2002, 08:52:12
 Job time: 5834 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 8, 2002, 08:05:20 ; Search time 2296.3 seconds
(without alignments)
1688.297 Million cell updates/sec

Title: US-09-823-101-2
Perfect score: 235
Sequence: 1 acctctaccactgctgac.....ttcttaggcaagtgaagctt 235

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl : *
1: gb_ba : *
2: gb_htg : *
3: gb_in : *
4: gb_om : *
5: gb_ov : *
6: gb_pat : *
7: gb_ph : *
8: gb_pl : *
9: gb_pr : *
10: gb_ro : *
11: gb_sts : *
12: gb_sy : *
13: gb_un : *
14: gb_vl : *
15: em_ba : *
16: em_fun : *
17: em_hum : *
18: em_in : *
19: em_om : *
20: em_or : *
21: em_ov : *
22: em_pat : *
23: em_ph : *
24: em_pl : *
25: em_ro : *
26: em_sts : *
27: em_sy : *
28: em_un : *
29: em_vl : *
30: em_htgo_hum : *
31: em_htgo_inv : *
32: em_htgo_rod : *
33: em_htg_hum : *
34: em_htg_inv : *
35: em_htg_rod : *
36: em_htg_other : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	43.8	18.6	7218	6	I66494	166494 Sequence 14
c 2	39.6	16.9	176697	2	AC021170	AC021170 Homo sapi
c 3	38.4	16.3	161261	9	CNS05TE3	AL358332 Human chr
c 4	38	16.2	150869	2	AC023992	AC023992 Homo sapi
c 5	38	16.2	158905	9	AC005666	AC005666 Homo sapi
c 6	37.8	16.1	123149	9	AC008655	AC008655 Homo sapi
c 7	37.8	16.1	198316	2	AC019157	AC019157 Homo sapi
c 8	37.4	15.9	153773	9	HUAC004020	AC004020 Homo sapi
c 9	37	15.7	161081	2	AC092015	AC092015 Homo sapi
c 10	37	15.7	161119	33	AL356368	AL356368 Human DNA
c 11	37	15.7	162518	2	AL592288	AL592288 Homo sapi
c 12	36.8	15.7	124288	2	AF207067	AF207067 Homo sapi
c 13	36.8	15.7	156833	9	AC018712	AC018712 Homo sapi
c 14	36.8	15.7	180192	2	AC020681	AC020681 Homo sapi
c 15	36.6	15.6	263	11	HUMUT5082A	L31021 Human STS U
c 16	36.6	15.6	889	5	AF089018	AF089018 Cacicus s
c 17	36.6	15.6	889	5	AF290169	AF290169 Cacicus s
c 18	36.6	15.6	77649	2	AC016654	AC016654 Homo sapi
c 19	36.6	15.6	112901	2	AP002511	AP002511 Homo sapi
c 20	36.6	15.6	152472	9	AP000486	AP000486 Homo sapi
c 21	36.6	15.6	161740	2	AC084778	AC084778 Homo sapi
c 22	36.6	15.6	164577	2	AC016549	AC016549 Homo sapi
c 23	36.6	15.6	171411	2	AC087609	AC087609 Homo sapi
c 24	36.6	15.6	177573	2	AC019076	AC019076 Homo sapi
c 25	36.6	15.6	177638	9	AC016652	AC016652 Homo sapi
c 26	36.6	15.6	178105	9	AC073057	AC073057 Homo sapi
c 27	36.6	15.6	203876	2	AC010413	AC010413 Homo sapi
c 28	36.2	15.4	141385	2	AC090714	AC090714 Oryza sat
c 29	36.2	15.4	170431	2	AC093331	AC093331 Homo sapi
c 30	36.2	15.4	172813	2	AC015720	AC015720 Homo sapi
c 31	36.2	15.4	174305	2	AC012191	AC012191 Homo sapi
c 32	36.2	15.4	196416	9	AC006483	AC006483 Homo sapi
c 33	36.2	15.4	251993	2	AC023822	AC023822 Homo sapi
c 34	36	15.3	2805	4	AF286474	AF286474 Bos tauru
c 35	36	15.3	117071	9	AL133375	AL133375 Human DNA
c 36	36	15.3	130526	9	AL133282	AL133282 Human DNA
c 37	36	15.3	163154	2	AC026532	AC026532 Homo sapi
c 38	36	15.3	166942	9	HS1054A22	AL031651 Human DNA
c 39	36	15.3	194405	2	AC032004	AC032004 Homo sapi
c 40	35.8	15.2	1968	14	PEN288589	AJ288589 Porcine e
c 41	35.8	15.2	158608	2	AC026834	AC026834 Homo sapi
c 42	35.8	15.2	220399	2	AC015918	AC015918 Homo sapi
c 43	35.6	15.1	115915	2	AP000859	AP000859 Homo sapi
c 44	35.6	15.1	153588	9	AC010533	AC010533 Homo sapi
c 45	35.6	15.1	161797	2	AP001775	AP001775 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS I66494/c
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dornier, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

28-DEC-1997

```

Query Match      18.6%; Score 43.8; DB 6; Length 7218;
Best Local Similarity 3.5%; Pred. No. 0.027;
Matches       6; Conservative    115; Mismatches     52; Indels      0; Gaps      0;

QY   19 acttggtgagtggggtccaaaggagttaggagataagaagtcaggctctcatagcca 78
Db   1443 ATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1384
          ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   79 atactatggtctacagttcccgaggaggaggagatccaaaccttgggaagaagacaagca 138
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   1383 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1324
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   139 gtagaagtgggcagtggggaggcaggaaaagtgtgggtcacagatcacaggaggagg 191
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   1323 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1271

RESULT 2
AC021170/c
LOCUS               AC021170             176697 bp            DNA                HTG
DEFINITION           Homo sapiens chromosome 11 clone RP11-655D7, WORKING DRAFT
                     SEQUENCE, 24 unordered pieces.
ACCESSION            AC021170
VERSION              AC021170.4 GI:7235310
KEYWORDS              HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE                human.
ORGANISM              Homo sapiens
                      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE             1 (bases 1 to 176697)
AUTHORS               Waterston,R.H.
TITLE                 The sequence of Homo sapiens clone
JOURNAL               Unpublished
REFERENCE             2 (bases 1 to 176697)
AUTHORS               Waterston,R.H.
TITLE                 Direct Submission
JOURNAL               Submitted (14-JAN-2000) Genome Sequencing Center, Washington
                      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                      MO 63108, USA
COMMENT               On Mar 13, 2000 this sequence version replaced gi:7024164.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0655D07
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164861 bases at least Q40
Consensus quality: 168822 bases at least Q30
Consensus quality: 170901 bases at least Q20
Insert size: 194000; agarose-fp
Quality coverage: 174397; sum-of-contigs
Quality coverage: 3.57 in Q20 bases; agarose-fp
Quality coverage: 4.02 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1086: contig of 1086 bp in length
* 1087 1186: gap of unknown length
* 1187 4671: contig of 3485 bp in length

```



```
repeat_region /rpt_family="L1MB3"
2471..2778
repeat_region /rpt_family="AluY"
2779..2925
repeat_region /rpt_family="AluJo/FRAM"
2926..2948
repeat_region /rpt_family="(CAAA)n"
2955..3255
repeat_region /rpt_family="L1MB3"
3256..3560
repeat_region /rpt_family="AluSc"
3570..3729
repeat_region /rpt_family="FAM"
3730..3787
repeat_region /rpt_family="L1MB3"
complement(3964..4114)
repeat_region /rpt_family="MER5B"
5111..5426
repeat_region /rpt_family="AluJb"
5480..5515
repeat_region /rpt_family="AT_rich"
5752..5784
repeat_region /rpt_family="AT_rich"
complement(5999..6328)
repeat_region /rpt_family="MER58B"
6435..6518
repeat_region /rpt_family="L1MB5"
6519..6815
repeat_region /rpt_family="AluY"
6816..7533
repeat_region /rpt_family="L1MB5"
7535..7694
repeat_region /rpt_family="AluJb"
7705..8001
repeat_region /rpt_family="AluSg"
8124..8183
repeat_region /rpt_family="MLT2D"
complement(8184..8205)
repeat_region /rpt_family="AluSg"
complement(8206..8229)
repeat_region /rpt_family="(CAAAA)n"
complement(8230..8515)
repeat_region /rpt_family="AluSg"
8516..8942
repeat_region /rpt_family="MLT2D"
complement(8957..9079)
repeat_region /rpt_family="MER7A"
9080..9419
repeat_region /rpt_family="THE1C"
complement(9420..9626)
repeat_region /rpt_family="MER7A"
9702..9748
repeat_region /rpt_family="MLT2D"
9863..9940
repeat_region /rpt_family="L2"
10022..10238
repeat_region /rpt_family="L1MB3"
complement(10325..11142)
repeat_region /rpt_family="L2"
complement(11143..11173)
repeat_region /rpt_family="(TCAA)n"
complement(11175..11440)
repeat_region /rpt_family="AluJb"
complement(11441..12055)
repeat_region /rpt_family="L2"
complement(12056..12364)
repeat_region /rpt_family="AluSg"
complement(12365..12447)
repeat_region /rpt_family="L2"
12448..12722
repeat_region /rpt_family="AluSc"
12726..13024
repeat_region /rpt_family="AluJb"
```

```
repeat_region complement(13025..13417)
/rpt_family="L2"
repeat_region 13419..13698
/rpt_family="AluSx"
repeat_region 13699..13726
/rpt_family="(TAAA)n"
repeat_region 13783..14076
/rpt_family="AluY"
repeat_region complement(14083..14322)
/rpt_family="MIR"
repeat_region 14484..14510
/rpt_family="AT_rich"
repeat_region complement(15900..16212)
/rpt_family="L1MA4A"
repeat_region 16280..16596
/rpt_family="MER7A"
repeat_region 16986..17306
/rpt_family="AluJo"
repeat_region 17422..17468
/rpt_family="(CAAAAA)n"
repeat_region 17550..17642
/rpt_family="HAL1"
repeat_region complement(17643..17937)
/rpt_family="AluSp"
repeat_region 17938..18133
/rpt_family="HAL1"
repeat_region 18764..19102
/rpt_family="L1ME"
repeat_region 19121..19742
/rpt_family="L1M4"
repeat_region 20318..20493
/rpt_family="MLT1J"
repeat_region complement(20498..20831)
/rpt_family="L1M4"
repeat_region complement(20888..21059)
/rpt_family="AluSg"
repeat_region complement(21060..21328)
/rpt_family="AluY"
repeat_region complement(21329..21460)
/rpt_family="AluSg"
repeat_region 22123..23454
/rpt_family="L1MB2"
repeat_region 23455..23586
/rpt_family="ELAM_C"
repeat_region 23587..24009
/rpt_family="L1MB2"
repeat_region 24010..24312
/rpt_family="AluSg"
repeat_region 24333..25000
/rpt_family="L1M1"
repeat_region 25001..25298
/rpt_family="AluSg"
repeat_region 25299..26025
/rpt_family="L1M1"
repeat_region 26026..26321
/rpt_family="AluSx"
repeat_region 26335..27573
/rpt_family="L1M1"
repeat_region 27574..27870
```

Query Match 16.2%; Score 38; DB 9; Length 158905;

Best Local Similarity 62.8%; Pred. No. 1.3;

Matches 59; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 124 gaagaagacaaggcagtagaagtggcgagtgaggagcaggaagggttggtgacagatc 183

Db 54330 GAAGATGATGATCAAGGGGAGGAAGAGAGAGATGAAGGAGAGATGACTAATAGAAC 54271

QY 184 agggagggtgtctgaccttttttttttggaggaaatt 217

Db 54270 ACTGATGGATTCTGACCTCTCTTTTAAAAAATTT 54237


```
* 32740 38306: contig of 5567 bp in length
* 38307 38406: gap of unknown length
* 38407 43714: contig of 5308 bp in length
* 43715 43814: gap of unknown length
* 43815 49339: contig of 5725 bp in length
* 49340 49639: gap of unknown length
* 49640 54932: contig of 5293 bp in length
* 54933 55032: gap of unknown length
* 55033 61252: contig of 6220 bp in length
* 61253 61352: gap of unknown length
* 61353 65754: contig of 4402 bp in length
* 65755 70345: gap of unknown length
* 70346 70445: contig of 4491 bp in length
* 70446 76104: gap of unknown length
* 76105 76204: contig of 5659 bp in length
* 76205 82987: gap of unknown length
* 82988 83087: contig of 6783 bp in length
* 83088 90305: contig of 7218 bp in length
* 90306 90406: gap of unknown length
* 90407 95481: contig of 5076 bp in length
* 95482 104523: contig of 8942 bp in length
* 104524 104624: gap of unknown length
* 104625 112124: contig of 7501 bp in length
* 112125 112225: gap of unknown length
* 112226 118912: contig of 6688 bp in length
* 118913 119012: gap of unknown length
* 119013 127446: contig of 8434 bp in length
* 127447 127546: gap of unknown length
* 127547 135381: contig of 7835 bp in length
* 135382 135481: gap of unknown length
* 135482 146537: contig of 11056 bp in length
* 146538 159888: gap of unknown length
* 159889 177512: contig of 17524 bp in length
* 177513 177612: gap of unknown length
* 177613 198316: contig of 20704 bp in length.
FEATURES             location/Qualifiers
    source
1..198316
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-25A12"
misc_feature
1..1530
/note="assembly_name:Contig8"
misc_feature
1631..2785
/note="assembly_name:Contig12"
misc_feature
2886..4726
/note="assembly_name:Contig15"
misc_feature
4827..6177
/note="assembly_name:Contig16"
misc_feature
6278..7608
/note="assembly_name:Contig17"
misc_feature
7709..10057
/note="assembly_name:Contig19"
misc_feature
10158..11576
/note="assembly_name:Contig20"
misc_feature
11677..12851
/note="assembly_name:Contig21"
misc_feature
12952..14468
/note="assembly_name:Contig22"
misc_feature
14569..16810
/note="assembly_name:Contig24"
misc_feature
16911..18701
/note="assembly_name:Contig25"
misc_feature
18802..21243
/note="assembly_name:Contig26"
misc_feature
21344..24777
/note="assembly_name:Contig27"
misc_feature
24878..28600
/note="assembly_name:Contig28"
```

```
misc_feature
28701..32639
/note="assembly_name:Contig29"
misc_feature
32740..38306
/note="assembly_name:Contig30"
misc_feature
38407..43714
/note="assembly_name:Contig31"
misc_feature
43815..49539
/note="assembly_name:Contig32"
misc_feature
49640..54932
/note="assembly_name:Contig33"
misc_feature
55033..61252
/note="assembly_name:Contig34"
misc_feature
61353..65754
/note="assembly_name:Contig35"
misc_feature
65855..70345
/note="assembly_name:Contig36"
misc_feature
70446..76104
/note="assembly_name:Contig37"
misc_feature
76205..82987
/note="assembly_name:Contig38"
misc_feature
83088..90305
/note="assembly_name:Contig39"
misc_feature
90406..95481
/note="assembly_name:Contig40"
misc_feature
95582..104523
/note="assembly_name:Contig41"
misc_feature
104624..112124
/note="assembly_name:Contig42"
misc_feature
112225..118912
/note="assembly_name:Contig43"
misc_feature
119013..127446
/note="assembly_name:Contig44"
misc_feature
127547..135381
/note="assembly_name:Contig45"
misc_feature
135482..146537
/note="assembly_name:Contig46"
misc_feature
146638..159888
/note="assembly_name:Contig47"
misc_feature
159889..177512
/note="assembly_name:Contig48"
misc_feature
177613..198316
/note="assembly_name:Contig49"
BASE COUNT 48035 a 50008 c 49087 g 47690 t 3496 others
ORIGIN
```

Query Match 16.1%; Score 37.8; DB 2; Length 198316;
Best Local Similarity 54.7%; Pred. No. 1.4;
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
QY 26 gagtgggtgtcccaaggaggtagagagataagagtcaggctctcatagccaataacta 85
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18100 GACGGGGGTAGAGAGGGTGTGGGGGGTCTGAGAAAAGGACGGAGAGAGAGAGAG 18159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 86 tggcttacagtcccgaggagggagatccaaacccctggaagaagacaaggcagtagaag 145
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18160 GGGGACAGGGTCCCCAGGAAGTGGAGACAGAGACCTAGAAAGAGAGAGAGAGAGAG 18219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 146 tgggcgagtgaggagca 162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18220 GTCAACATAGGAGGAA 18236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 8
HUAC004020/c      HUAC004020 153773 bp      DNA      PRI      23-NOV-1999
LOCUS      Homo sapiens Chromosome 16 BAC clone CIT987SK-A-319E8, complete
DEFINITION      sequence.
ACCESSION      AC004020.
VERSION      AC004020.1 GI:3219330
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
```


Siddiqui, R., Taudien, S., Wen, G., Korenberg, J. R., Rosenthal, A. and Platzer, M.

TITLE
Chromosome 8 genomic sequenceUnpublished
2 (bases 1 to 124288)REFERENCE
Reichwald, K., Wen, G., Schillhabel, M., Baumgart, C., Dette, M., Menzel, U., Jahn, N. and Rosenthal, A.TITLE
Direct SubmissionJOURNAL
Submitted (19-NOV-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
On Jun 2, 2001 this sequence version replaced gi:8152098.

COMMENT

----- Genome Center
Center: Institute of Molecular Biotechnology

Center code: IMB

Web site: <http://genome.imb-jena.de/>Contact: gscj-submit@genome.imb-jena.de

----- Project Information

Center project name: H208

Center clone name: XX-CTB875C8

----- Summary Statistics

Sequencing vector: M13; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 117740 bases at least Q40

Consensus quality: 120559 bases at least Q30

Consensus quality: 123333 bases at least Q20

Quality coverage: 5.69 x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.* 1 4113: contig of 4113 bp in length
* 4114 4213: gap of unknown length
* 4214 47267: contig of 43054 bp in length
* 47268 47367: gap of unknown length
* 47368 124288: contig of 76921 bp in length.
* Location/Qualifiers
1. .124288
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="Bq24.1"
/clone="XX-CTB875C8"
1. .4
/note="assembly_fragment
clone_end:SP6
vector_side:left"
124285. .124288
/note="assembly_fragment
clone_end:T7
vector_side:right"

FEATURES

source

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>Contact: sapiens@watson.wustl.edu

----- Summary Statistics

----- Center project name: H_NH0343H10

misc_feature

vector_side:left

/note="assembly_fragment

clone_end:T7

vector_side:right

BASE COUNT 36681 a 26600 c 25566 g 35241 t 200 others

ORIGIN

Query Match

Best Local Similarity 15.7%; Score 36.8; DB 2; Length 124288;

Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 100 gaggaaggagatccaaacccctgaaagaagacagcagtagaagtgagcagtgaggag 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14865 GAGGAGGAGGAGACATGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 14806
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 160 gcagaaagggttgatcacagatcaggagggtgtctgaccttttcttgaggaaattct 219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14805 GCAGAAAGGAGAGAGACAGACAGAAAGAGTCTACCATTTCTCGAGCTCGATGTAT 14746
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 220 taggcaag 227
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14745 TAGGTAAG 14738
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13

AC018712/c

LOCUS AC018712 156833 bp DNA PRI 09-MAY-2001

DEFINITION Homo sapiens BAC clone RP11-343H10 from 2, complete sequence.

ACCESSION AC018712

VERSION AC018712.5 GI:11415210

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 156833)

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

REFERENCE 2 (bases 1 to 156833)

Edwards, J., Maupin, R., Hawkins, M. and Smith, R.

The sequence of Homo sapiens BAC clone RP11-343H10

AUTHORS Unpublished

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 156833),

Waterston, R.H.

Direct Submission

Submitted (16-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE 4 (bases 1 to 156833)

Waterston, R.H.

Direct Submission

Submitted (29-NOV-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE 5 (bases 1 to 156833)

Waterston, R.

Direct Submission

Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Nov 29, 2000 this sequence version replaced gi:9838147.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>Contact: sapiens@watson.wustl.edu

----- Summary Statistics

----- Center project name: H_NH0343H10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACes.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-313011, 200 bp overlap; the clone sequenced to the right is AC020681. Actual start of this clone is at base position 198520 of RP11-313011; actual end is at base position 156833 of RP11-343H10.

Bases 123072 to 123647 of RP11-343H10 consists only of a PCR product from clone DNA. Bases 62320 to 62338 consists only of a single M13 subclone. Digest and PCR information confirms the assembly.

FEATURES

.source	Location/Qualifiers
	1. .156833
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-343H10"
	/clone_lib="RPCI-11"
repeat_region	937. .1272
repeat_region	/rpt_family="L2"
repeat_region	1280. .1408
repeat_region	/rpt_family="L2"
repeat_region	1639. .2020
repeat_region	/rpt_family="L1"
repeat_region	2021. .2389
repeat_region	/rpt_family="MaLR"
repeat_region	3971. .4333
repeat_region	/rpt_family="MaLR"
repeat_region	4334. .4849
repeat_region	/rpt_family="L1"
repeat_region	4850. .5138
repeat_region	/rpt_family="Alu"
repeat_region	5139. .5421
repeat_region	/rpt_family="L1"
repeat_region	5700. .6000
repeat_region	/rpt_family="Alu"
repeat_region	6914. .8085
repeat_region	/rpt_family="L1"
repeat_region	8086. .8143
repeat_region	/rpt_family="MIR"
repeat_region	8172. .8324
repeat_region	/rpt_family="MIR"
repeat_region	8798. .8978
repeat_region	/rpt_family="MaLR"
repeat_region	10040. .10639
repeat_region	/rpt_family="L1"
repeat_region	10640. .10945
repeat_region	/rpt_family="Alu"
repeat_region	10946. .10995
repeat_region	/rpt_family="L1"
repeat_region	11259. .11317
repeat_region	/rpt_family="L2"
repeat_region	12438. .12762
repeat_region	/rpt_family="MERL_type"
repeat_region	12942. .13248
repeat_region	/rpt_family="Alu"
repeat_region	14289. .14321
repeat_region	/rpt_family="L2"
repeat_region	18392. .18494
repeat_region	/rpt_family="MIR"
repeat_region	19330. .19607
repeat_region	/rpt_family="L1"
repeat_region	19668. .19778
repeat_region	/rpt_family="L1"
repeat_region	19794. .19930
repeat_region	/rpt_family="Alu"
repeat_region	20395. .20617
repeat_region	/rpt_family="MIR"
repeat_region	21510. .21934
repeat_region	/note="similar to
repeat_region	EST AV6955578 (NID:g10297441)"
repeat_region	21510. .22135
repeat_region	/note="similar to
repeat_region	EST AV684851 (NID:g10286714)"
repeat_region	21510. .22057
repeat_region	/note="similar to
repeat_region	EST AV691567 (NID:g10293430)"
repeat_region	21861. .22339
repeat_region	/rpt_family="ERV1"
repeat_region	22106. .22228
repeat_region	/note="similar to
repeat_region	EST AV691567 (NID:g10293430)"
repeat_region	22968. .23304
repeat_region	/rpt_family="ERV1"
repeat_region	23305. .23608
repeat_region	/rpt_family="Alu"
repeat_region	23609. .23740
repeat_region	/rpt_family="ERV1"
repeat_region	25588. .25870
repeat_region	/rpt_family="L2"
repeat_region	26252. .26701
repeat_region	/rpt_family="MaLR"
repeat_region	26774. .27034
repeat_region	/rpt_family="L2"
repeat_region	27040. .27583
repeat_region	/rpt_family="L1"
repeat_region	27593. .27801
repeat_region	/rpt_family="L1"
repeat_region	27836. .28064
repeat_region	/rpt_family="MaLR"
repeat_region	28403. .28493
repeat_region	/rpt_family="L1"
repeat_region	28654. .28768
repeat_region	/rpt_family="L1"
repeat_region	28909. .29145
repeat_region	/rpt_family="L1"
repeat_region	29939. .30091
repeat_region	/rpt_family="MERL_type"
repeat_region	30032. .30522
repeat_region	/note="similar to
repeat_region	EST AL120277 (NID:g5926176)"
repeat_region	30271. .30730
repeat_region	/rpt_family="MaLR"
repeat_region	30975. .31269
repeat_region	/rpt_family="Alu"
repeat_region	31281. .31769
repeat_region	/rpt_family="L1"
repeat_region	31880. .32055
repeat_region	/rpt_family="L1"
repeat_region	32089. .32193
repeat_region	/rpt_family="L1"
repeat_region	32198. .32495
repeat_region	/rpt_family="Alu"
repeat_region	32534. .32943
repeat_region	/rpt_family="L1"
repeat_region	33070. .33734
repeat_region	/rpt_family="L1"
repeat_region	34622. .34704
repeat_region	/rpt_family="MERL_type"
repeat_region	34700. .34860
repeat_region	/rpt_family="MERL_type"

Query Match 15.7%; Score 36.8; DB 9; Length 156833;
Best Local Similarity 58.0%; Pred. No. 2.8;

```
Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
Qy 110 agatccaaacctggaagacagagcagtagaagtggcgagtggaagcaggaag 169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51633 ACAGGGAGACCTGCAAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGG 51574
Qy 170 ttgggtgacatcagggaggggtgctgtaccttttttcttgaggaaattctta 221
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51573 AGAGGGAGGAGAGGAGGAGGAGGAGGAGCTTTTCTTAATTAATTTTA 51522

RESULT 14
AC020681/c
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-366L15 map 2, WORKING DRAFT
ACCESSION AC020681
VERSION AC020681.3 GI:9280773
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 180192)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-366L15
Unpublished
2 (bases 1 to 180192)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczkyl,J., Levine,R., Lien,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Menees,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisan,R.C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 19, 2000 this sequence version replaced gi:6731248.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIPR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3778
Center clone name: 366_L15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173966 bases at least Q40
Consensus quality: 176894 bases at least Q30
Consensus quality: 178103 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 178992; sum-of-contigs
Quality coverage: 5.1 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
```

* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1617: contig of 1617 bp in length
1618 1717: gap of 100 bp
1718 6392: contig of 4675 bp in length
6393 6492: gap of 100 bp
6493 11804: contig of 5312 bp in length
11805 11904: gap of 100 bp
11905 18137: contig of 6233 bp in length
18138 18237: gap of 100 bp
18238 25025: contig of 6788 bp in length
25026 25125: gap of 100 bp
25126 33095: contig of 7970 bp in length
33096 33195: gap of 100 bp
33196 46602: contig of 13407 bp in length
46603 46702: gap of 100 bp
46703 61609: contig of 14907 bp in length
61610 61709: gap of 100 bp
61710 80467: contig of 18758 bp in length
80468 80567: gap of 100 bp
80568 97584: contig of 17017 bp in length
97585 97684: gap of 100 bp
97685 119038: contig of 21354 bp in length
119039 119138: gap of 100 bp
119139 148789: contig of 29651 bp in length
148790 148889: gap of 100 bp
148890 180192: contig of 31303 bp in length.

Location/Qualifiers
1..180192
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-366L15"
/clone_lib="RPC1-11 Human Male BAC"
1..1617
/note="assembly_fragment"
1718..6392
/note="assembly_fragment"
6493..11804
/note="assembly_fragment"
11905..18137
/note="assembly_fragment"
clone_end:SP6
vector_side:left
18238..25025
/note="assembly_fragment"
25126..33095
/note="assembly_fragment"
clone_end:T7
vector_side:right
33196..46602
/note="assembly_fragment"
46703..61609
/note="assembly_fragment"
61710..80467
/note="assembly_fragment"
80568..97584
/note="assembly_fragment"
97685..119038
/note="assembly_fragment"
119139..148789
/note="assembly_fragment"
148890..180192
/note="assembly_fragment"
BASE COUNT 53893 a 36593 c 36153 g 52350 t 1203 others
ORIGIN

THIS PAGE BLANK (USPTO)